REMARKS

Claims 1, 2, 7-13, and 18-22 are presently pending in the application. Claims 1, 2, 7-13 and 18-22 have been amended and Claims 3-6 and 14-17 have been cancelled without prejudice. Applicants wish to thank the Examiner for providing a copy of Menasce et al. "Capacity Planning for Web Performance, Metrics, Models and Methods," 1998 ("CWPA(98)") with the Office Action dated February 26, 2004. In view of the above claim amendments and arguments presented hereinbelow, Applicants respectfully submit that these claims are now in condition for allowance.

Claim Rejections -- 35 U.S.C. § 103(a)

Claims 1-22 stand rejected under Section 103(a) as being unpatentable over ("CWPA(98)") in view of Li et al. U.S. Patent No. 5,583,792 ("Li") or Waclawsky et al. U.S. Patent No. 5,197,127 ("Waclawsky"). Applicants respectfully traverse the rejection and submit that the combination of these references fails to teach or suggest the claimed invention.

The present invention provides a method and apparatus for modeling a web server to enable the performance of Web server platforms to be evaluated and compared, by modeling a server as a plurality of subsystems. In an illustrative embodiment, a model in accordance with an aspect of the invention captures key performance limiting factors of an HTTP server. As described in the specification:

> The results lead to fast-to-evaluate approximations of the key performance measures of a Web server, like the server throughput, the end-to-end service time, and the blocking probabilities. Page 4, line 23 – Page 5, line 3.
>
> . . . .
>
> The embodiments of the invention are useful for modeling and analysis of the impact of the different components on Web server performance. Accordingly, the embodiments of the invention describe a new analytic model for Web servers. HTTP transactions proceed along a number of phases in successive order. Therefore, the HTTP flows within a Web server can be described

by a tandem queuing model consisting of the following sub-models:

- a multi-server, zero-buffer blocking model for the TCP/IP connection set-up phase;
- a multi-server, finite-buffer queuing/blocking model for the HTTP server processing; and
- a multi-server, infinite-buffer queuing model for the network I/O sub-system. Page 5, lines 7 – 17.

As set forth in claim 1, as amended, an exemplary method for modeling a web server in accordance with the invention, comprises the steps of:

> *identifying a plurality of sub-systems* for the server, said plurality of sub-systems *comprising a transaction control protocol/internet protocol (TCP/IP) sub-system, a hypertext transfer protocol (HTTP) sub-system, and an input/output (IO) sub-system*;
>
> representing each sub-system as a queue, with each queue operably coupled together, *wherein said TCP/IP sub-system comprises a first finite listen queue served by a listener daemon, said HTTP sub-system comprises a second finite listen queue served by one or more multi-threaded HTTP daemons with N_{http} separate server threads, and said I/O sub-system comprises a finite number N_{buf} of network buffers served by an input/output controller*; and
>
> iteratively adjusting an arrival rate and a service time for each queue to account for performance by other queues. Emphasis added.

None of the cited art, either taken alone or in combination, teaches or suggests the invention of claim 1.

The Examiner contends that CPWP (98) "is explicit in teaching models for constructing queuing systems that cover the web servers and other internet systems that interact with web servers. . . .However [CPWP(98)] does not go into great detail as to combining the metrics with these models for interactively and iteratively improving the network server operations as is clearly delineated in Li et al. and Waclawsky et al., individually." Office Action dated August 27, 2003 at Page 4, lines 5 – 12. However, the Examiner does not assert that CPWP(98) teaches modeling a web server by sub-models represented by queuing systems.

Applicants agree that CPWP(98) generally discloses the notion of a queuing network as an amalgamation of interconnected queues that represent a computer system. However, Applicants respectfully disagree that the combination of CPWP(98) and Li or Waclawsky discloses or suggests the claimed invention. With respect to Web servers, CPWP(98) describes a queuing model that is based on HTTP requests. As described in CPWP(98) at Chapter 10, page 238:

> The queuing network model corresponding to Fig. 10.5a is shown in Fig. 10.6a. We are assuming here that we are dealing with a Web server that is publicly available on the Internet. Thus, there is a very large population of unknown size of clients that will access the Web server. Thus, we can only characterize the arrival rate of requests for various document sizes. Therefore, we will model the Web server as an open multiclass QN model. Different classes in the model correspond to HTTP requests of different size as discussed in Ex. 10.3. Let *R* be the number of classes and λ_r (r=1, . . . , *R*) be the arrival rate of class r requests. As explained in Ex. 10.3, the average arrival rate is computed as
>
> $$\lambda_r = \lambda \text{ x PercentSize}_r$$
>
> where λ is the overall arrival rate of HTTP requests to the Web server.
>
> As in Sec. 10.3, the incoming and outgoing links and the LAN are represented by load-impendent queues and the router as a delay queue. The Web server is represented by two load-independent queues: one for the CPU and another for the disk. One could have many more disks and a multiprocessor CPU as well.

CPWP(98) discloses that a Web server may be modeled by representing HTTP requests as an open multiclass queuing model. However, CPWP(98) contains absolutely no teaching or suggestion of modeling a Web server by establishing a TCP/IP subsystem, a HTTP subsystem *and* an I/O subsystem as claimed. Moreover, CPWP(98) fails to disclose or suggest the claimed steps of "representing each sub-system as a queue, with each queue operably coupled together, wherein said *TCP/IP sub-system comprises a first finite listen queue served by a listener daemon*, said *HTTP sub-system comprises a second finite listen queue served by one or more multi-threaded HTTP daemons with* N_{http}

separate server threads, and said *I/O sub-system comprises a finite number N_{buf} of network buffers served by an input/output controller*." CPWP(98) only discloses the use of an open multiclass QN model where different classes in the model correspond to HTTP requests of different sizes. This does not correspond to a finite listen queue as claimed. Moreover, this disclosure fails to suggest the additional steps of modeling a TCP/IP subsystem and an I/O subsystem in a Web server queuing model. The addition of Li and Waclawsky fails to remedy the deficiencies in the disclosure of CPWP(98).

Turning first to Li, that reference contains absolutely no teaching or suggestion of modeling a web server by "representing each sub-system as a queue, with each queue operably coupled together" and "iteratively adjusting an arrival rate and a service time for each queue to account for performance by other queues." Moreover, Li fails to disclose or suggest the specifically claimed steps of "said TCP/IP sub-system comprises a first finite listen queue served by a listener daemon, said HTTP sub-system comprises a second finite listen queue served by one or more multi-threaded HTTP daemons with N_{http} separate server threads, and said I/O sub-system comprises a finite number N_{buf} of network buffers served by an input/output controller." Li is concerned with an entirely different issue, namely, a method and apparatus which provides a general solution technique for the integration of traffic measurement and queuing analysis. See Abstract. Applicants are unable to find anything in Li that suggests modeling a server in the particular manner that is claimed by applicants. In fact, Li doesn't even mention anything about Web servers. This is acknowledged by the Examiner in the Office Action dated August 27, 2003 at page 4. Thus, even if, assuming *arguendo*, CPWP(98) and Li are properly combinable, such combination would fail to reach the claimed invention.

With respect to Waclawsky, that patent discloses a method of data flows in a communications network. As described in Waclawsky:

> The method includes the steps of setting a packet transmission window to have a maximum quantity of N packets which can be transmitted within an interval from a terminal in the network and setting a queued packet threshold value to a quantity of C packets

> which may be held in a queue during an interval at the terminal. The method can also be applied to an intermediate node in the network.
>
> The method then defines a data flow efficiency variable S as a binary number having at least three bits, with a first bit B1 which assumes a binary value of one if the number of packets transmitted by the terminal during an interval is equal to N, a second bit B2 which assumes a binary value of one if any packet is held in the queue during an interval and a third bit B3 which assumes a value of one if more than C packets are held in the queue during an interval.
>
> The method then counts the number of packets transmitted from the terminal during a measurement period and sets B1 equal to one if the number of packets transmitted in any interval during the period is equal to N, it sets B2 equal to one if any packet is held in the queue during the measurement period, and it sets B3 equal to one if more than C packets are held in the queue during any interval in the measurement period.
>
> The method then determines the value of the data flow efficiency state variable S from values of B1, B2 and B3 set by the counting and setting steps and it accesses a knowledge base containing network problem determination recommendations which are accessible with the value of the data flow efficiency state variable S.
>
> Finally, the method outputs a problem determination recommendation for optimizing data flow efficiency in the network in response to accessing the knowledge base with the value of S. Col. 1, line 67 – Col. 2, line 35.

There is nothing in Waclawsky that even remotely discloses or suggests modeling a server using the claimed steps of "representing each sub-system as a queue, with each queue operably coupled together" and "iteratively adjusting an arrival rate and a service time for each queue to account for performance by other queues" as called for in independent claim 1 (and those claims dependent on claim 1). To the contrary, Waclawsky teaches that a node in a network, for example, node 1, may be represented as a pair of queues 17, 20. The queue 17 is a buffer that holds surplus packets at node 1 that are unable to be transmitted during a particular interval. See Col. 3, lines 36 – 44. Queue 20 also holds surplus packets that have

been accumulated during a particular transmission interval. See Col. 3, line 65 – Col. 4, line 2. This disclosure clearly does not teach or suggest the claimed steps discussed above as none of these queues have anything to do with the subsystems of a Web server as set forth above. Moreover, this disclosure fails to disclose or suggest the specific algorithms set forth in the dependent claims. Accordingly, even if, assuming *arguendo*, CPWP(98) and Waclawsky are properly combinable, such combination still doesn't reach the claimed invention as Waclawsky fails to remedy the deficiencies in the disclosure of CPWP(98). The same analysis applies to independent claims 11 and 22, as well as those claims dependent thereon.

In view of the above, Applicants respectfully submit that claims 1, 2, 7-13, and 18-22 are patentable over the cited art, and allowance of these claims at an early date is solicited.

The Office is hereby authorized to charge any additional fees or credit any overpayments under 37 C.F.R. 1.16 or 1.17 to AT&T Corp. Account No. 01-2745. The Examiner is invited to contact the undersigned at (201) 224-7957 to discuss any matter concerning this application.

Respectfully submitted,
Paul Kevin Reeser, et al.
By:

Date: 9/22/04

Gary H. Monka
Registration No. 35,290
Attorney for Applicant

Canavan & Monka, LLC.
805 Partridge Drive
Bridgewater, New Jersey 08807
(201) 224-7957